

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

THIS PAGE BLANK (USPTO)



(19) Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) Publication number:

0 679 005 A1

(12)

EUROPEAN PATENT APPLICATION

(21) Application number: 95104367.8

(51) Int. Cl.⁶: H04M 3/50

(22) Date of filing: 24.03.95

(30) Priority: 22.04.94 GB 9408042

(43) Date of publication of application:
25.10.95 Bulletin 95/43

(64) Designated Contracting States:
DE FR GB

(71) Applicant: Hewlett-Packard Company
3000 Hanover Street
Palo Alto,
California 94304 (US)

(72) Inventor: Haddock, Nicholas John
23 Buchanans Wharf South,
Ferry Street
Bristol BS1 6HJ (GB)

(74) Representative: Webster, Denise Mary et al
Hewlett-Packard Ltd,
IP Section,
Building 2,
Filton Road,
Stoke Gifford
Bristol BS12 6QZ (GB)

(54) Device for managing voice data.

(57) The present invention concerns a device for managing voice data. The embodiment described comprises means (20) for displaying a visual representation of a voice message and means for associating markers (42,44,46,48) with segments of the message. The markers (42,44,46,48) are indicative of particular storage areas eg a telephone number storage area, a calendar storage area etc. Association of a marker (42,44,46,48) with a segment of a voice message automatically causes that segment to be linked with the corresponding storage area so that the segment can later be retrieved in the context of a user interface for that particular storage area.

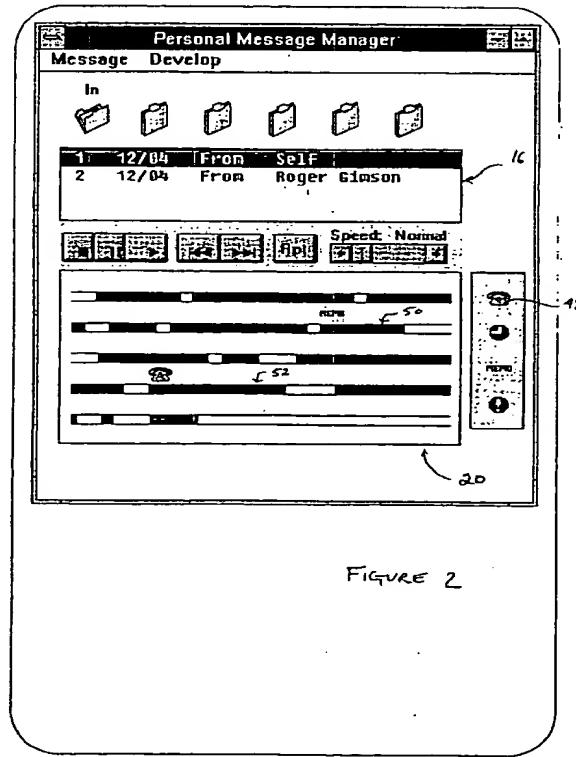


FIGURE 2

Technical Field

The present invention relates to a device designed to facilitate the management of voice data. Voice messages, left on a recipient's answerphone or delivered via a voicemail system are a popular form of person-to-person communication. Such voice messages are quick to generate for the sender but are relatively difficult to review for the recipient; speech is slow to listen to and, unlike inherently visual forms of messages such as electronic mail or handwritten notes, cannot be quickly scanned for the relevant information. The present invention aims to make it easier for users to extract relevant information from voice messages, and other kinds of voice record, such as recordings of meetings and recorded dictation.

In the long-term it would be desirable to approach this problem by automatically translating speech into text using speech recognition. Unfortunately this approach is not yet practical, since current speech recognition technology cannot accurately transcribe naturally-occurring speech of the kind found in voice messages. Therefore a number of approaches have been developed which help users to review voice data without actually recognising the speech signal and which provide for the display, structuring and annotation of speech recordings.

Background Art

Many approaches assume, but do not necessarily depend on, an underlying technique for displaying a visual representation of speech. One such form of display is a single graphical line, graduated with time markings from start to finish (for example, a 4 second message may contain the appropriately spaced labels "0 sec", "1 sec", "2 sec", "3 sec", "4 sec"). In addition, an algorithm can be used to process the speech record to distinguish the major portions of speech from the major portions of silence. Such an algorithm is described by Arons (1994, Chapter 4). This permits a richer form of graphical display, in which the speech record is still portrayed along a timeline, but with portions of speech displayed as dark segments (for example) and the detected portions of silence displayed as light segments. Four pieces of prior art will be referred to:

1. A paper in the proceedings of CHI '92 entitled "Working with Audio: Integrating Personal Tape Recorders and Desktop Computers" by Degen, Mander and Saloman (1992) describes a prototype hand-held personal tape recorder. This is similar to a conventional "dictaphone" except that the user can place index points on the recording by pressing a button at the appro-

priate point in the recording. Two index buttons are available and these have no predetermined meaning. The user is free to place their own interpretation on the two forms of index. The recording can be downloaded to a personal computer and the inserted index points can be displayed along the timeline of the message. By visually displaying the index points, the user is reminded of an area of interest in the speech recording and can selectively play back portions of speech by using a pointing device such as a mouse. In addition, the index points can be searched for within the recording.

2. The NoteTaker product from InkWare Development Corp. (1994) extends this idea in the context of computer-based handwritten notes, rather than speech. Here users can select one of a variety of visual labels, representing for example "Urgent!", "Call" or "Action", and associate these with selected parts of a handwritten note. The program then allows the user to find all notes containing a particular label, an "Action" item for example.

3. Ades and Swinehart (1986) have built a prototype system for annotating and editing speech records. This system is the subject of their paper entitled "Voice Annotation and Editing in a Workstation Environment" from Xerox Corporation. In particular, an arbitrary text annotation can be placed on a visually displayed segment of speech as a cue to the content of that portion of speech.

4. A paper entitled "Capturing, Structuring and Representing Ubiquitous Audio" by Hindus, Schmandt and Horner (ACM Transactions on Information Systems, Vol 11, No.4 October 1993, pages 376-400) describes a prototype system for handling speech which allows the user to select a portion of visually displayed speech and to associate the depicted speech portion (such as by "drag-and-drop" using a mouse) with another application, such as a calendar. The calendar may contain independently entered, standard textual data (such as "Meeting with Jim"), as well as audio annotations and additions associated in this way.

Referring to the prior art items numbered 1-4 above, approaches (1) - (3) offer annotations which the user can employ as a visual cue to relevant parts of the speech (or handwriting, in the case of (2)). In (1), two labels are available with no predefined meaning. In (2), the user can choose from a broader set of labels, the appearance of which suggests a particular use (eg. the user should use the "Call" label for tagging items about telephoning people). In (3), the user can tag speech with an arbitrary textual entry, thus providing an even richer form of annotation. However, in

all these approaches the label plays only a passive role in organising the target data. It is a passive visual and searchable cue to parts of the speech, and does not help the broader integration of the speech with other relevant applications in the user's personal information environment.

Approach (4) addresses this problem by allowing users to associate selected speech clips into, for example, a text-based calendar. A disadvantage of this approach is that it is rather laborious - the user must identify the appropriate speech clip, select it, and then associate it with another application. In addition, not all user interfaces lend themselves to this approach.

Disclosure of Invention

According to the present invention we provide a device for storing speech input comprising: means for specifying a marker having a particular connotation; means for associating the marker with all or part of the speech input; and means for automatically linking the speech input associated with the marker to a corresponding storage area for later retrieval by the user in the context of a user interface which is dependent on the connotation of the associated marker.

A device according to the present invention has the advantage of providing a simple and convenient way of integrating voice data with other user applications so as to facilitate the management of voice data. In the embodiment to be described, the corresponding storage areas include telephone book and calendar application storage areas.

Preferably, the means for specifying a marker comprises means for selecting a marker from a set of markers. The set of markers preferably comprise iconic representations of the corresponding storage areas.

In the embodiment to be described there are means for displaying a representation of the speech input. This allows a user to view a visual representation of voice data on a desktop computer display. In that embodiment, there are means for automatically segmenting the speech input, specifically for automatically segmenting the speech input into silent and non-silent parts.

The marker may be associated with a part of the speech input by time synchronisation. This approach conflates the selection of a marker and its association with a segment of speech data in a manner which may be particularly convenient for users. Alternatively, the marker may be associated with a part of the speech input by user input. The user input may comprise manipulation of an input device eg. dragging and dropping a marker icon on

the relevant speech segment using a mouse. Alternatively, the user input may comprise means for associating a marker with a part of the speech input by spoken commands.

5 The linking means may comprise means for copying the speech input associated with the marker to the corresponding storage area. Alternatively, the linking means may comprise means for moving the speech input associated with the marker to the 10 corresponding storage area. Another possibility is for the linking means to comprise means for providing a pointer to the speech input associated with the marker in the corresponding storage area. It may also be useful for the linking means to comprise means for providing an index into the original 15 voice data containing the speech input associated with the marker.

Brief Description of Drawings

20 Particular embodiments of the present invention will now be described, by way of example, with reference to the accompanying drawings of which:

25 Figure 1 depicts the user interface of a device according to a first embodiment of the present invention;
Figure 2 depicts the user interface of Figure 1 after labelling of two speech segments;
Figure 3 depicts the user interface of a known 30 telephone book application.

Best Mode for Carrying Out the Invention & Industrial Applicability

35 The present invention can be implemented in the context of a "Personal Message Manager" application for browsing voice messages.

The embodiment to be described with reference to Figures 1 to 3 is written in Microsoft Visual 40 Basic and Borland C on a IBM-compatible 486 25MHz Personal Computer, and runs under the Microsoft Windows 3.1 operating system. Audio recording and playback facilities are supported by a SoundBlaster 16ASP card (Creative Labs, Inc.). These facilities are accessed through the standard 45 MS Windows MultiMedia Application Programmers' Interface. Speech records are created using a microphone connected to the audio card, and played back via a set of speakers also connected to the card. On recording, the audio card translates the 50 analogue audio signal produced by the microphone into a standard digital representation of the recorded speech, and stores the data in the standard ".wav" file format. The card performs the converse digital-to-analogue conversion in order to play back 55 a digital ".wav" file through loudspeakers.

User input is by means of a mouse.

Figure 1 shows an interaction screen 10 in a Microsoft Windows user interface. A set of folders represented by icons 12 are for storing previous voice messages. One of the folders 14 has been selected which causes the "header" information for each message in the selected folder to be displayed in a display box 16. The display box 16 displays the date of receipt and the sender of each message. Figure 1 shows the topmost message 18 having been selected. This causes the selected message 18 to be displayed as a series of blocks in another display box 20. In the display box 20, dark blocks represent speech and white blocks represent silence. A known speech processing algorithm is utilised to distinguish between the major segments of speech and silence; such an algorithm is described in the paper by Arons (1994, Chapter 4).

Above the display box 20 is a set of audio controls 22 to allow the user to play, pause and stop speech playback. The audio controls 22 comprise the following button representations:

- a play button 24;
- a pause button 26;
- a stop button 28;
- a previous button 30 to skip playback to the previous segment of speech;
- a next button 32 to skip playback to the next segment of speech;
- a repeat button 34 to repeat playback of the most recently played segment of speech;
- a speed control button 36 to vary the playback speed.

The user can also click directly on a segment of speech in the display box 20 eg using a mouse, to play back that specific segment. In Figure 1, an arrow-shaped cursor 38 is shown in the display box 20 to indicate that playback is ready to commence at the beginning of the speech file. As a speech segment is being played, its colour changes to give the user a cue to the current position in the speech record.

To the right of the display box 20 is a panel 40 of markers 42,44,46, and 48 for labelling portions of the recorded speech. These can be used to provide a visual cue to the contents of a message. There are markers corresponding to a Phone Book 42, a Time/Appointment diary 44, a Memo/Reminder list 46, and a miscellaneous Points of Interest area 48. For example, one segment of the message 18 may contain a speech segment such as "If you need to get back to me, my number is 228 455." This segment could be labelled with the Phone marker 42. Whenever a marker is placed on a speech segment in the display box 20, that segment of speech is automatically linked to a corresponding application in the user's computer system. This automatic linking

of speech segments to other applications using visual markers is convenient for the user and is an important step towards integrating the various applications relevant to handling voice data.

5 Figure 2 depicts a situation in which the user has labelled two segments of speech, 50 and 52, the segment 50 as a Memo, and the segment 52 as a Phone item. This is accomplished by clicking the appropriate marker during playback of the relevant speech segment; the system then associates an instance of this marker with the segment of speech being played and provides a visual representation of the marker above the segment in the display box 20 as shown.

10 15 As well as providing a visual cue to the content of the speech record, placing markers against speech segments in the display box 20 automatically links the labelled segments to an appropriate computer application. For example, marking the message with the Phone label 42 as shown in Figure 2 causes the marked segment of speech to be automatically added to a standard, textual Phone Book application, depicted in Figure 3. The 'Phone Book' window comprises a display box 54 listing the entries in the directory and two buttons,

20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080 1085 1090 1095 1100 1105 1110 1115 1120 1125 1130 1135 1140 1145 1150 1155 1160 1165 1170 1175 1180 1185 1190 1195 1200 1205 1210 1215 1220 1225 1230 1235 1240 1245 1250 1255 1260 1265 1270 1275 1280 1285 1290 1295 1300 1305 1310 1315 1320 1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 1395 1400 1405 1410 1415 1420 1425 1430 1435 1440 1445 1450 1455 1460 1465 1470 1475 1480 1485 1490 1495 1500 1505 1510 1515 1520 1525 1530 1535 1540 1545 1550 1555 1560 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 1615 1620 1625 1630 1635 1640 1645 1650 1655 1660 1665 1670 1675 1680 1685 1690 1695 1700 1705 1710 1715 1720 1725 1730 1735 1740 1745 1750 1755 1760 1765 1770 1775 1780 1785 1790 1795 1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865 1870 1875 1880 1885 1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050 2055 2060 2065 2070 2075 2080 2085 2090 2095 2100 2105 2110 2115 2120 2125 2130 2135 2140 2145 2150 2155 2160 2165 2170 2175 2180 2185 2190 2195 2200 2205 2210 2215 2220 2225 2230 2235 2240 2245 2250 2255 2260 2265 2270 2275 2280 2285 2290 2295 2300 2305 2310 2315 2320 2325 2330 2335 2340 2345 2350 2355 2360 2365 2370 2375 2380 2385 2390 2395 2400 2405 2410 2415 2420 2425 2430 2435 2440 2445 2450 2455 2460 2465 2470 2475 2480 2485 2490 2495 2500 2505 2510 2515 2520 2525 2530 2535 2540 2545 2550 2555 2560 2565 2570 2575 2580 2585 2590 2595 2600 2605 2610 2615 2620 2625 2630 2635 2640 2645 2650 2655 2660 2665 2670 2675 2680 2685 2690 2695 2700 2705 2710 2715 2720 2725 2730 2735 2740 2745 2750 2755 2760 2765 2770 2775 2780 2785 2790 2795 2800 2805 2810 2815 2820 2825 2830 2835 2840 2845 2850 2855 2860 2865 2870 2875 2880 2885 2890 2895 2900 2905 2910 2915 2920 2925 2930 2935 2940 2945 2950 2955 2960 2965 2970 2975 2980 2985 2990 2995 3000 3005 3010 3015 3020 3025 3030 3035 3040 3045 3050 3055 3060 3065 3070 3075 3080 3085 3090 3095 3100 3105 3110 3115 3120 3125 3130 3135 3140 3145 3150 3155 3160 3165 3170 3175 3180 3185 3190 3195 3200 3205 3210 3215 3220 3225 3230 3235 3240 3245 3250 3255 3260 3265 3270 3275 3280 3285 3290 3295 3300 3305 3310 3315 3320 3325 3330 3335 3340 3345 3350 3355 3360 3365 3370 3375 3380 3385 3390 3395 3400 3405 3410 3415 3420 3425 3430 3435 3440 3445 3450 3455 3460 3465 3470 3475 3480 3485 3490 3495 3500 3505 3510 3515 3520 3525 3530 3535 3540 3545 3550 3555 3560 3565 3570 3575 3580 3585 3590 3595 3600 3605 3610 3615 3620 3625 3630 3635 3640 3645 3650 3655 3660 3665 3670 3675 3680 3685 3690 3695 3700 3705 3710 3715 3720 3725 3730 3735 3740 3745 3750 3755 3760 3765 3770 3775 3780 3785 3790 3795 3800 3805 3810 3815 3820 3825 3830 3835 3840 3845 3850 3855 3860 3865 3870 3875 3880 3885 3890 3895 3900 3905 3910 3915 3920 3925 3930 3935 3940 3945 3950 3955 3960 3965 3970 3975 3980 3985 3990 3995 4000 4005 4010 4015 4020 4025 4030 4035 4040 4045 4050 4055 4060 4065 4070 4075 4080 4085 4090 4095 4100 4105 4110 4115 4120 4125 4130 4135 4140 4145 4150 4155 4160 4165 4170 4175 4180 4185 4190 4195 4200 4205 4210 4215 4220 4225 4230 4235 4240 4245 4250 4255 4260 4265 4270 4275 4280 4285 4290 4295 4300 4305 4310 4315 4320 4325 4330 4335 4340 4345 4350 4355 4360 4365 4370 4375 4380 4385 4390 4395 4400 4405 4410 4415 4420 4425 4430 4435 4440 4445 4450 4455 4460 4465 4470 4475 4480 4485 4490 4495 4500 4505 4510 4515 4520 4525 4530 4535 4540 4545 4550 4555 4560 4565 4570 4575 4580 4585 4590 4595 4600 4605 4610 4615 4620 4625 4630 4635 4640 4645 4650 4655 4660 4665 4670 4675 4680 4685 4690 4695 4700 4705 4710 4715 4720 4725 4730 4735 4740 4745 4750 4755 4760 4765 4770 4775 4780 4785 4790 4795 4800 4805 4810 4815 4820 4825 4830 4835 4840 4845 4850 4855 4860 4865 4870 4875 4880 4885 4890 4895 4900 4905 4910 4915 4920 4925 4930 4935 4940 4945 4950 4955 4960 4965 4970 4975 4980 4985 4990 4995 5000 5005 5010 5015 5020 5025 5030 5035 5040 5045 5050 5055 5060 5065 5070 5075 5080 5085 5090 5095 5100 5105 5110 5115 5120 5125 5130 5135 5140 5145 5150 5155 5160 5165 5170 5175 5180 5185 5190 5195 5200 5205 5210 5215 5220 5225 5230 5235 5240 5245 5250 5255 5260 5265 5270 5275 5280 5285 5290 5295 5300 5305 5310 5315 5320 5325 5330 5335 5340 5345 5350 5355 5360 5365 5370 5375 5380 5385 5390 5395 5400 5405 5410 5415 5420 5425 5430 5435 5440 5445 5450 5455 5460 5465 5470 5475 5480 5485 5490 5495 5500 5505 5510 5515 5520 5525 5530 5535 5540 5545 5550 5555 5560 5565 5570 5575 5580 5585 5590 5595 5600 5605 5610 5615 5620 5625 5630 5635 5640 5645 5650 5655 5660 5665 5670 5675 5680 5685 5690 5695 5700 5705 5710 5715 5720 5725 5730 5735 5740 5745 5750 5755 5760 5765 5770 5775 5780 5785 5790 5795 5800 5805 5810 5815 5820 5825 5830 5835 5840 5845 5850 5855 5860 5865 5870 5875 5880 5885 5890 5895 5900 5905 5910 5915 5920 5925 5930 5935 5940 5945 5950 5955 5960 5965 5970 5975 5980 5985 5990 5995 6000 6005 6010 6015 6020 6025 6030 6035 6040 6045 6050 6055 6060 6065 6070 6075 6080 6085 6090 6095 6100 6105 6110 6115 6120 6125 6130 6135 6140 6145 6150 6155 6160 6165 6170 6175 6180 6185 6190 6195 6200 6205 6210 6215 6220 6225 6230 6235 6240 6245 6250 6255 6260 6265 6270 6275 6280 6285 6290 6295 6300 6305 6310 6315 6320 6325 6330 6335 6340 6345 6350 6355 6360 6365 6370 6375 6380 6385 6390 6395 6400 6405 6410 6415 6420 6425 6430 6435 6440 6445 6450 6455 6460 6465 6470 6475 6480 6485 6490 6495 6500 6505 6510 6515 6520 6525 6530 6535 6540 6545 6550 6555 6560 6565 6570 6575 6580 6585 6590 6595 6600 6605 6610 6615 6620 6625 6630 6635 6640 6645 6650 6655 6660 6665 6670 6675 6680 6685 6690 6695 6700 6705 6710 6715 6720 6725 6730 6735 6740 6745 6750 6755 6760 6765 6770 6775 6780 6785 6790 6795 6800 6805 6810 6815 6820 6825 6830 6835 6840 6845 6850 6855 6860 6865 6870 6875 6880 6885 6890 6895 6900 6905 6910 6915 6920 6925 6930 6935 6940 6945 6950 6955 6960 6965 6970 6975 6980 6985 6990 6995 7000 7005 7010 7015 7020 7025 7030 7035 7040 7045 7050 7055 7060 7065 7070 7075 7080 7085 7090 7095 7100 7105 7110 7115 7120 7125 7130 7135 7140 7145 7150 7155 7160 7165 7170 7175 7180 7185 7190 7195 7200 7205 7210 7215 7220 7225 7230 7235 7240 7245 7250 7255 7260 7265 7270 7275 7280 7285 7290 7295 7300 7305 7310 7315 7320 7325 7330 7335 7340 7345 7350 7355 7360 7365 7370 7375 7380 7385 7390 7395 7400 7405 7410 7415 7420 7425 7430 7435 7440 7445 7450 7455 7460 7465 7470 7475 7480 7485 7490 7495 7500 7505 7510 7515 7520 7525 7530 7535 7540 7545 7550 7555 7560 7565 7570 7575 7580 7585 7590 7595 7600 7605 7610 7615 7620 7625 7630 7635 7640 7645 7650 7655 7660 7665 7670 7675 7680 7685 7690 7695 7700 7705 7710 7715 7720 7725 7730 7735 7740 7745 7750 7755 7760 7765 7770 7775 7780 7785 7790 7795 7800 7805 7810 7815 7820 7825 7830 7835 7840 7845 7850 7855 7860 7865 7870 7875 7880 7885 7890 7895 7900 7905 7910 7915 7920 7925 7930 7935 7940 7945 7950 7955 7960 7965 7970 7975 7980 7985 7990 7995 8000 8005 8010 8015 8020 8025 8030 8035 8040 8045 8050 8055 8060 8065 8070 8075 8080 8085 8090 8095 8100 8105 8110 8115 8120 8125 8130 8135 8140 8145 8150 8155 8160 8165 8170 8175 8180 8185 8190 8195 8200 8205 8210 8215 8220 8225 8230 8235 8240 8245 8250 8255 8260 8265 8270 8275 8280 8285 8290 8295 8300 8305 8310 8315 8320 8325 8330 8335 8340 8345 8350 8355 8360 8365 8370 8375 8380 8385 8390 8395 8400 8405 8410 8415 8420 8425 8430 8435 8440 8445 8450 8455 8460 8465 8470 8475 8480 8485 8490 8495 8500 8505 8510 8515 8520 8525 8530 8535 8540 8545 8550 8555 8560 8565 8570 8575 8580 8585 8590 8595 8600 8605 8610 8615 8620 8625 8630 8635 8640 8645 8650 8655 8660 8665 8670 8675 8680 8685 8690 8695 8700 8705 8710 8715 8720 8725 8730 8735 8740 8745 8750 8755 8760 8765 8770 8775 8780 8785 8790 8795 8800 8805 8810 8815 8820 8825 8830 8835 8840 8845 8850 8855 8860 8865 8870 8875 8880 8885 8890 8895 8900 8905 8910 8915 8920 8925 8930 8935 8940 8945 8950 8955 8960 8965 8970 8975 8980 8985 8990 8995 9000 9005 9010 9015 9020 9025 9030 9035 9040 9045 9050 9055 9060 9065 9070 9075 9080 9085 9090 9095 9100 9105 9110 9115 9120 9125 9130 9135 9140 9145 9150 9155 9160 9165 9170 9175 9180 9185 9190 9195 9200 9205 9210 9215 9220 9225 9230 9235 9240 9245 9250 9255 9260 9265 9270 9275 9280 9285 9290 9295 9300 9305 9310 9315 9320 9325 9330 9335 9340 9345 9350 9355 9360 9365 9370 9375 9380 9385 9390 9395 9400 9405 9410 9415 9420 9425 9430 9435 9440 9445 9450 9455 9460 9465 9470 9475 9480 9485 9490 9495 9500 9505 9510 9515 9520 9525 9530 9535 9540 9545 9550 9555 9560 9565 9570 9575 9580 9585 9590 9595 9600 9605 9610 9615 9620 9625 9630 9635 9640 9645 9650 9655 9660 9665 9670 9675 9680 9685 9690 9695 9700 9705 9710 9715 9720 9725 9730 9735 9740 9745 9750 9755 9760 9765 9770 9775 9780 9785 9790 9795 9800 9805 9810 9815 9820 9825 9830 9835 9840 9845 9850 9855 9860 9865 9870 9875 9880 9885 9890 9895 9900 9905 9910 9915 9920 9925 9930 9935 9940 9945 9950 9955 9960 9965 9970 9975 9980 9985 9990 9995 9999

the time is noted say, 5324ms, and then the speech/silence data file, illustrated above, is searched to see to which segment this time corresponds. In the above example this implies that the system is currently playing the third speech segment. In this way, time-synchronization is used to associate a marker with a speech segment.

In order automatically to link to another application and subsequently to play a speech clip from that application, a visual indication of the speech within that application is provided and the relevant application must be able to play back the speech clip directly. This is accomplished using standard MS Windows programming techniques. In the Phone Book example, an automatically generated textual entry is added to the Phone Book display (for example, see item 60 in Figure 3). In addition, in the underlying data structure, this entry is flagged as being voice data and a simple specification of where to find the appropriate voice data is recorded. This specification comprises a pointer to the original ".wav" speech file, along with a specification of start and end points within this file that represent the speech segment to be accessed. These points can be specified as times, byte positions, or other representations. When selected, the audio Application Programmers' Interface is used to play back this segment of speech from within the Phone Book application.

The embodiment described above is a voice data management device which is easy to use and which integrates voice data into other user applications in a convenient manner. Many of the features described with reference to this embodiment can be modified and categories of these will now be addressed.

1. Selection of speech marker

Apart from a mouse, other possible selection devices include a pen/stylus, a touch-screen and the use of the TAB key on a keyboard for iteratively cycling through menu selection options displayed to the user. Alternatively, each marker could be represented by a dedicated hard button on a device implementing the present invention and pressed during playback of recorded speech.

2. Association of markers with speech

In the embodiment described above, the timing of the marker selection governs the speech segment with which it is to be associated. An alternative is to allow the user actively to associate a marker with the speech segment of interest eg by "drag-and-drop". This approach is particularly useful after the message has been listened to at least once when the user is undertaking considered ana-

lysis and structuring of the speech file.

An alternative set of approaches conflate the selection and association steps. The user may select the speech segment of interest, either by explicit selection with a mouse, or implicit selection by time synchronisation, and linguistically specify the marker to be associated with that segment. The linguistic specification could be made by typing in some initial identifying characters of the name of the marker (eg. "ph" for Phone), by drawing or hand-writing the name of the marker and using handwriting recognition to determine the intended marker, or by speaking the name of the marker and using speech recognition to identify it.

5
10
15
20
25

A final general approach to marker association is automatically to identify the appropriate marker for a segment of speech by partially recognising the speech itself. Here techniques for "word-spotting" in continuous speech, for example based on Hidden Markov Models (cf. Wilcox and Bush, 1991), could determine the likelihood that a certain speech segment contains a telephone number. If the recognition algorithm predicts a high probability of a phone number, the segment could be labelled automatically with the Phone marker.

3. Definition of markers

It is anticipated that the user may be able to 30
35
40
45

customise the markers and corresponding storage areas available within a system according to the present invention. A suite of icons could be made available from which the user can choose. In addition, the user could define arbitrary text labels and place these in the panel of markers. The system could also allow the user to specify the storage area associated with each marker.

4. Accessing speech segments from target application

The above description assumes that the storage application (eg. Phone Book) is provided with a 45
50
55

link to the original speech file. There are various ways in which this could be implemented:

- i) Copy - a copy of the appropriate speech data could be made and stored in a separate file;
- ii) Move - a copy of the appropriate speech data could be made and stored in a separate file, and the segment could be removed from the original voice record (ie. from the voice message);
- iii) Link - as in the above-described embodiment, a pointer to the same speech file can be provided.

Another approach is to treat the copied/linked speech clip as representing an index into the original message. In this case, when the clip is played back from the application (eg. the Phone Book), the

user has the option of reviewing the entire message from which it was extracted. This is a useful enhancement since an automatic segmentation algorithm will sometimes produce inappropriate segmentations, for example breaking a telephone number in the middle, in which case it is important for the user to be able to continue playback (or rewind) after the linked speech clip has been played.

5. Extracting segments from the original speech record

In the above-described embodiment, the speech record is segmented into speech and silence using an algorithm such as Arons (1994, Chapter 4). Alternatively, the original speech record could be represented to the user as a continuous, unstructured line. Markers could be associated with this line using the same range of techniques described above and the only difference would be that the marker is associated with a point in the speech record rather than a segment of speech.

Automatically storing the speech associated with a marker could then be accomplished by either (a) arbitrarily defining the segment of interest eg. a 5 second clip centred on the marker point, or (b) assuming the indexing approach outlined in point (4) above, where the storage of the speech in the target application is merely a point at which to index into the original.

6. User interface designs

Whenever a marker is associated with a segment of speech, an instance of that marker could appear in the "header" line for the message (along with date, sender, etc). This would provide a cue to the user that the message contains eg. a phone number. A possible additional feature would be to play back every segment in the relevant message which has been associated with this type of marker on selection of the header marker by the user eg by clicking with the mouse.

Moreover, a "find" facility could be included with the Personal Message Manager which could find all messages containing a certain type of marker, or combination of marker.

7. Device without a display

The present invention also has application in a device which lacks a display. Such a device may be useful for visually impaired people, for whom speech-based information is more useful than visual information. Speech messages could be reviewed using a set of hard buttons, similar to those used in dictaphones for example, and interesting portions of speech could be labelled using a set of

hard marker buttons (as described in (1) above). Such portions could then be linked as described above to speech-based storage areas, such as a speech-based phone book.

5 The present invention is relevant to a range of uses of speech data. It may have particular utility for users who receive a large amount of voice mail containing similar kinds of information. This information may not need to be transcribed immediately, but it may help to store the spoken information in a structured form. For example, field staff may telephone a central office to report the time of a repair, the problem diagnosed and the work undertaken. This information could be extracted from the voice messages and categorised using the techniques described.

10 20 25 The invention has been described in terms of a program for handling voice messages. However, the invention is applicable to all forms of recorded speech, and the implementation described need not necessarily be part of a telecommunications system. Other possible uses include the management of voice data comprising recording of meetings, general conversations and other personal data.

Claims

- 30 1. A device for storing speech input comprising:
means for specifying a marker having a particular connotation;
means for associating the marker with all or part of the speech input;
and means for automatically linking the speech input associated with the marker to a corresponding storage area for later retrieval by the user in the context of a user interface which is dependent on the connotation of the associated marker.
- 35 40 2. A device according to claim 1 wherein the means for specifying a marker comprises means for selecting a marker from a set of markers.
- 45 3. A device according to claim 2 wherein the set of markers comprises iconic representations of the corresponding storage areas.
- 50 4. A device according to any preceding claim comprising means for displaying a representation of the speech input.
- 55 5. A device according to claim 4 comprising means for automatically segmenting the speech input.

6. A device according to claim 5 comprising means for automatically segmenting the speech input into silent and non-silent parts.
7. A device according to any preceding claim comprising means for associating a marker with a part of the speech input by time synchronisation. 5
8. A device according to any of claims 1 to 6 comprising means for associating a marker with a part of the speech input by user input. 10
9. A device according to claim 8 comprising means for associating a marker with a part of the speech input by manipulation of an input device. 15
10. A device according to claim 9 comprising means for associating a marker with a part of the speech input by spoken commands. 20
11. A device according to any preceding claim wherein the linking means comprises means for copying the speech input associated with the marker to the corresponding storage area. 25
12. A device according to any one of claims 1 to 10 wherein the linking means comprises means for moving the speech input associated with the marker to the corresponding storage area. 30
13. A device according to any one of claims 1 to 10 wherein the linking means comprises means for providing a pointer to the speech input associated with the marker in the corresponding storage area. 35
14. A device according to any one of claims 1 to 10 wherein the linking means comprises means for providing an index into the original voice data containing the speech input associated with the marker. 40

45

50

55

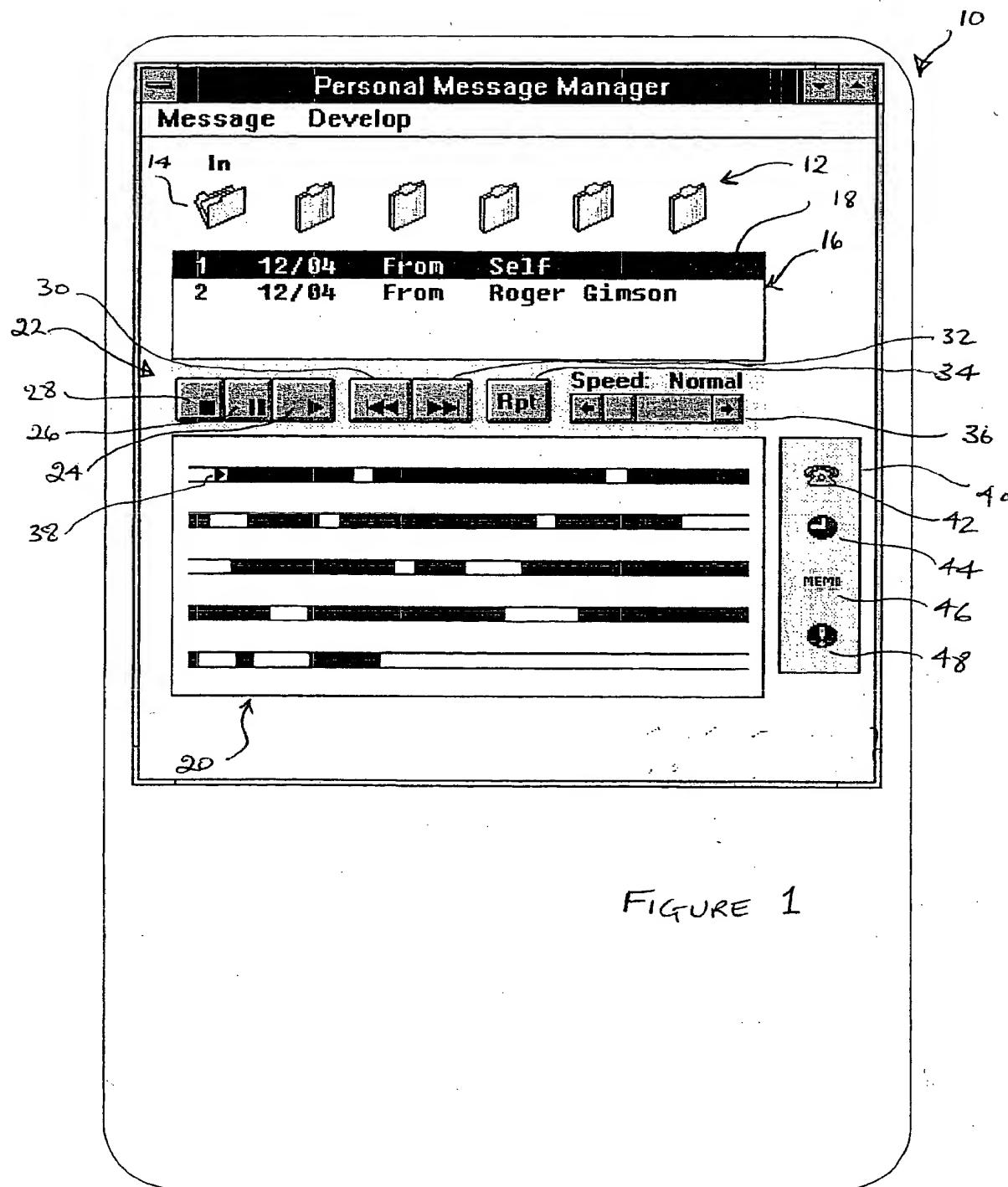
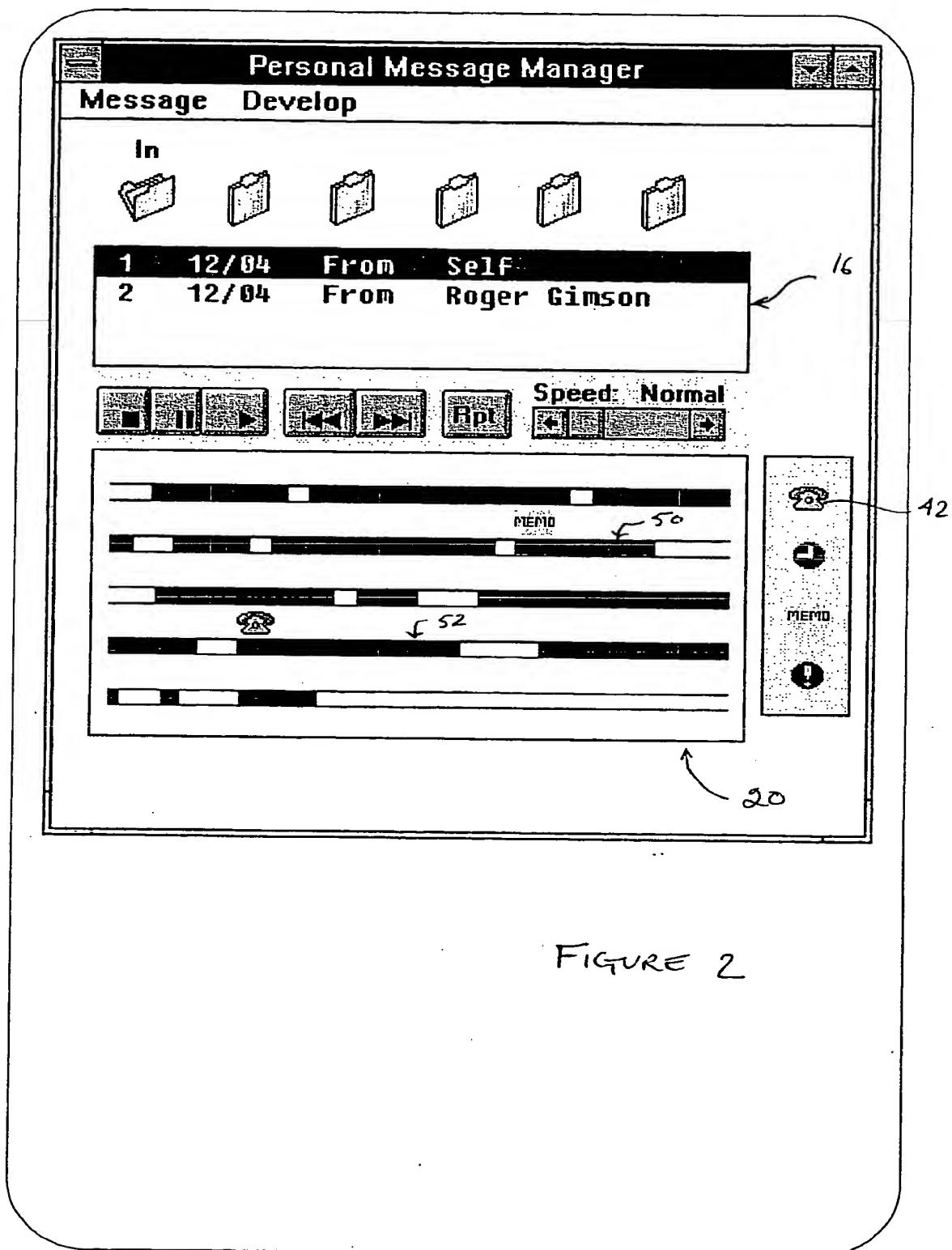


FIGURE 1



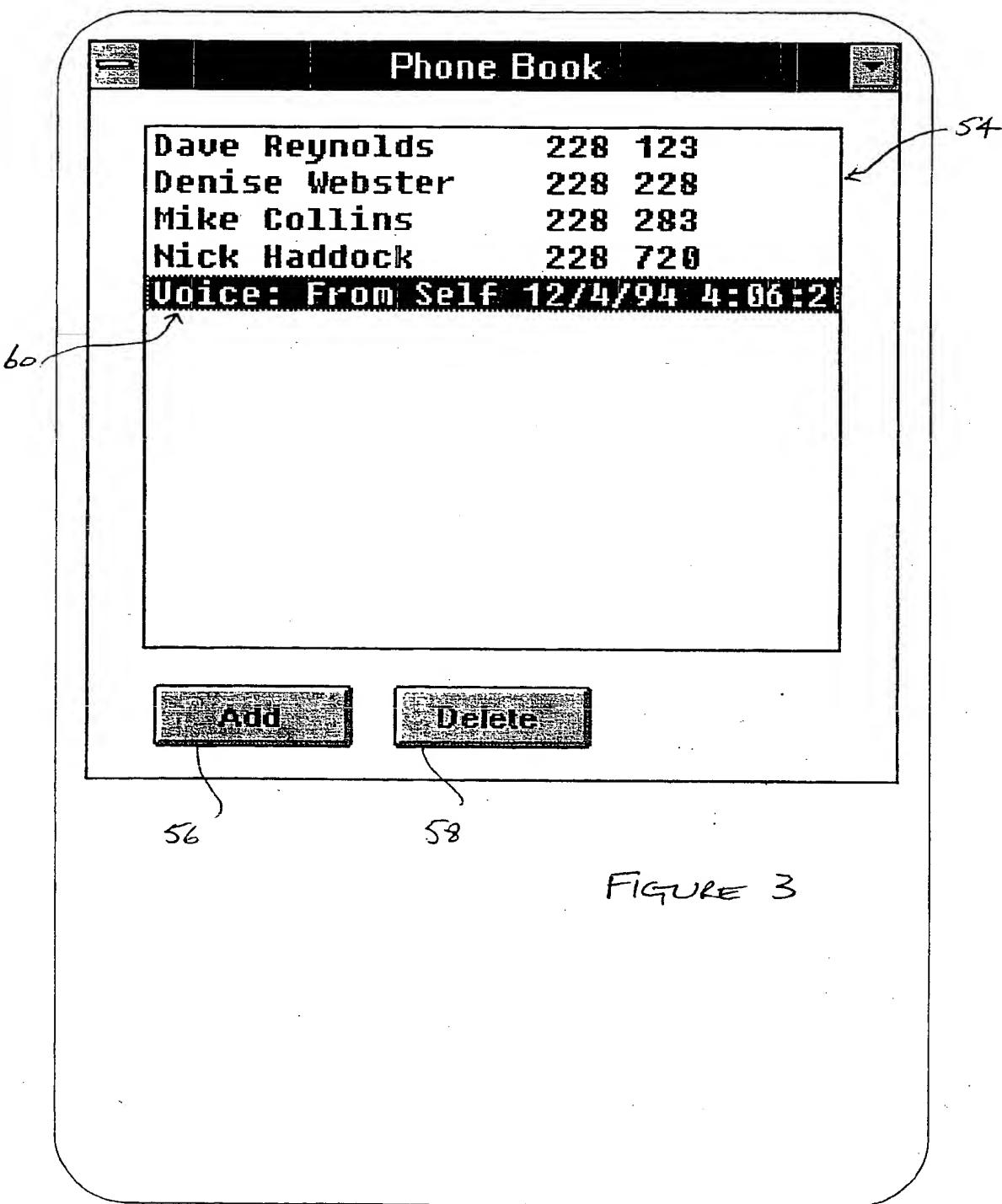


FIGURE 3



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 95 10 4367

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.)
X	PROCEEDINGS OF THE CONFERENCE ON COMPUTER-SUPPORTED COOPERATIVE WORK, 31 October 1992, TORONTO(CA) pages 210 - 217 DEBBY HINDUS ET AL 'UBIQUITOUS AUDIO : CAPTURING SPONTANEOUS COLLABORATION' * page 211, left column, line 18 - line 29 * Y * page 212, right column, line 24 - page 213, right column, line 21 * * page 215, left column, line 5 - line 14 * D, Y CHI 92- CONF. PROC. ACM CONF ON HUMAN FACTORS IN COMPUTER SYSTEMS, 3 May 1992, MONTEREY(US) pages 413 - 418 LEO DEGEN ET AL 'WORKING WITH AUDIO : INTEGRATING PERSONAL TAPE RECORDERS AND DESKTOP COMPUTERS' * page 415, left column, line 3 - right column, line 2 *	1, 4-9, 11-14 2, 3 2	H04M3/50
Y	WO-A-92 02009 (YOUNGER) * abstract *	2, 3	TECHNICAL FIELDS SEARCHED (Int.Cl.)
A	IBM TECHNICAL DISCLOSURE BULLETIN, vol.36, no.9B, September 1993, NEW YORK US pages 297 - 299 'METHOD OF CATEGORIZING CONCATENATED PHONE MESSAGES INTO MESSAGES LOGS' * the whole document *	-/--	H04M G10L G06F
The present search report has been drawn up for all claims			
Place of search	Date of completion of the search	Examiner	
THE HAGUE	14 June 1995	Vandevenne, M	
CATEGORY OF CITED DOCUMENTS		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document	
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document			



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 95 10 4367

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
A	IBM TECHNICAL DISCLOSURE BULLETIN, vol.36, no.6B, June 1993, NEW YORK US pages 169 - 170 'CATEGORICAL STORAGE OF VOICE MAIL MESSAGES' * the whole document * ---		
A	PROC. OF THE 1990 CONFERENCE- THE AMERICAN VOICE I/O SOCIETY, 1990, SAN JOSE (US) pages 71 - 75 CHRIS SCHMANDT 'CALTALK : A MULTI-MEDIA CALENDAR' ---		
A	WO-A-92 11634 (ARDIS) * page 4, line 28 - page 5, line 30 * -----		
The present search report has been drawn up for all claims			
Place of search THE HAGUE	Date of completion of the search 14 June 1995	Examiner Vandevenne, M	
CATEGORY OF CITED DOCUMENTS			
X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document		T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons R : member of the same patent family, corresponding document	